DOCUMENT RESUME

ED 441 118 CE 080 129

TITLE Feedback Systems. Symposium 28. [Concurrent Symposium Session at AHRD Annual Conference, 2000.]
PUB DATE 2000-03-08
NOTE 27p.; In: Academy of Human Resource Development Conference Proceedings (Raleigh-Durham, North Carolina, March 8-12, 2000); see CE 080 095.
PUB TYPE Collected Works - General (020) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Adults; Business Administration; Educational Needs; Evaluation Methods; *Feedback; Foreign Countries; Human Resources; *Labor Force Development; Learning Strategies; *Management Development; *Measurement Techniques; *Organizational Development; Program Effectiveness; Program Evaluation; *Staff Development
IDENTIFIERS United Kingdom; *360 Degree Feedback

ABSTRACT

This packet contains three papers from a symposium on feedback systems held at a conference on human resource development (HRD). The first paper, "The Role of Feedback in Management Development Training" (K. Peter Kuchinke), reports on a survey-based study that investigated the role of feedback in nine management development training settings in a British government agency. The results of the study suggest that participants sought information about their performance frequently and from a variety of sources and that the feedback-seeking is important in the process of management development training. The second paper, "A Five Phase Framework for Designing a Successful Multirater Feedback System" (Allan H. Church, Janine Waclawski) notes that there is a need in the literature for more attention to the factors involved in creating a successful feedback process. A five-phase framework for designing such a system is based on the classic organizational development consulting skills model and many years of practitioner experience with large-scale feedback-based applications in Fortune 100 organizations. The final paper, "An Evaluation of the Quality of 360-Degree Assessment Instruments" (Froukje Jellema, Adrie Visscher, Martin Mulder), presents a checklist of standards on the basis of which 360-degree employee evaluation instruments are examined. Four 360-degree assessment instruments also are examined. The papers contain reference sections.(KC)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 441 118

2000 AHRD Conference

Feedback Systems

Symposium 28

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Rochelle Chute
Jarlene
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Raleigh-Durham, NC

March 8 - 12, 2000

BEST COPY AVAILABLE

CE080129

The Role of Feedback in Management Development Training

K. Peter Kuchinke
University of Illinois at Urbana-Champaign

This survey-based study investigated the role of feedback in nine management development training settings in a British government agency. Distinguishing among different sources and types of feedback as provided by the instructors and sought by the participant, the results of this study suggest that participants sought information about their performance frequently and from a variety of sources. Instructors tended to overestimate the importance of the feedback they provided. The amount of feedback sought was related to judgements of relevance of the training and on the teaching styles employed by the instructors. The study indicates the feedback seeking is important in the process of management development training. The implications of these findings for further research and the practice of management development are discussed.

Keywords: Management Development, Feedback, International HRD

Management training and development ranks among the most frequently provided types of training. Research results of a nation-wide study in the United Kingdom in 1999 (Institute of Personnel and Development, 1999) revealed that among 400 randomly selected private and public organizations it ranked first with over 75% of organizations providing "a lot" of management training and an additional 20% providing "some". In the United States, it ranks second in frequency after new employee orientation with 93% of companies providing this kind of training (Bassie & Van Buren, 1998). Management training and development (MD) is broadly defined as "the attempt to improve managerial effectiveness through a planned and deliberate learning process" (de Bettignies, 1975, p. 4); the two most important goals of MD programs, according to a survey by the Conference Board are to develop leadership skills in managers and to insure a pool of capable people to run the organization (Walter, 1996). MD typically includes training in areas such as performance appraisals, implementing regulations and policies, managing projects and processes, and planning and budgeting (Bassie & Van Buren, 1998) and is directed to broad range of employees ranging from first-line supervisors and team leaders to mid-level managers. MD is distinct from executive development which is usually targeted towards current and potential senior executives and focuses on corporation-wide initiatives or major business units and includes strategic planning, policy making, and goal setting (Bassie & Van Buren, 1998).

The demand for MD has been attributed to the flattening of organizations and the broadening of job roles at many levels. Leadership behavior is no longer seen as the domain of a select few at the top of the organizational hierarchy but is required at all levels. Almost any individual in an organization, as Van de Ven and Grazman (1995) observed, "may act as a leader ... .Because the sharing of influence increases the quality of the decisions and the motivation of organizational participants, [it is proposed that] the more influential acts (i.e.leadership) are widely shared in an organization, the more effective the organization" (p.7).

While much has been written about the importance of management and leadership skills (for the purpose of this paper, both terms will be used interchangeably, conceptual discussions over their differences notwithstanding), there is less research about the content of MD and even less about the process of MD. This paper will report the results of an empirical study conducted in a series of MD settings focusing on one key instructional process element, feedback.

Feedback is a classic concept in the social sciences and has been the focus of research in fields ranging from control theory and cybernetics to psychology and education. Feedback has been defined as a " special case of the general communication process in which some sender... conveys a message... to a recipient"(Ilgen, Fisher, & Taylor, 1979, p. 350) related to some aspect of the recipient's behavior. Feedback is seen as critical for the functioning, maintenance, and adaptation of individuals, groups, or organizations because it provides information about the system's current state compared to some standard (Taylor, Fisher, & Ilgen, 1984). Without feedback, goal attainment would be improbable and action haphazard and random.

Feedback is a key component of any learning process. Successful training programs incorporate feedback as an instructional design element (Goldstein, 1993; Kovitz & Smith, 1985) and also during instructional delivery to

Copyright © 2000, K. Peter Kuchinke

increase learning and the transfer of learning (Schoenfeldt, 1996). Recent research articles have addressed the role of feedback in different training and education settings, for instance in industry training (Viau & Clark, 1987), for supervisors providing in-service staff training (Parsons & Reid, 1995), and in college and university education settings (Dunkins & Precians, 1992, Brinko, 1993).

The focus of this study was a particular aspect of feedback, the process of feedback seeking by participants of MD programs. Feedback seeking has been described as process by which actors purposefully and actively seek to obtain information to "determine the adequacy of behaviors for attaining valued end states" (Ashford, 1986, p. 466). In a comprehensive review of the literature related to feedback seeking, Madzar (1995) asserted the importance of the concept for HRD practice and suggested its important role for training in general and management development in particular.

Feedback seeking research is based on the premise that employees continually engage in a process of seeking information about two aspects of their behavior: which goals to pursue (how to direct their energy) and what progress they make toward these goals. They seek this information frequently, in a variety of ways, and from a variety of sources. Whereas traditional feedback research has treated the receiver of feedback in a relatively passive role, researchers in feedback seeking ascribe a very active role to the individual and "toss the ball into the seeker's court. That is, they denied the assumption that individuals are passive recipients of feedback ... and proposed that individuals actively seek feedback in order to have more control over the outcomes of their behavior" (Madzar, 1995, p. 337).

This current study was built upon the assumption that individuals self-regulate to a large extent (Bandura, 1986) and are actively involved in seeking information to monitor their progress towards specific goals. Management development programs are especially well suited to investigate feedback seeking because they present novel situations for participants who are advancing in an organization. New behaviors, knowledge, and skills are introduced which are of importance to employees who will assume new levels of responsibility. MD often serves as a rite of initiation and signals impending enhanced status and responsibilities. One key dimension of leadership is what Conover (1987, p. 585) termed "managing self" which includes monitoring progress towards goals and evaluating one's skills, strengths, and weaknesses. Feedback seeking behavior is an important source of information with regard to this dimension.

Research Questions and Review of the Literature

The study addresses four overall research questions:

(1) What information sources did participants use when seeking feedback on their performance?,
(2) What individual antecedents affected feedback seeking?
(3) What instructor-related antecedents affected feedback seeking?, and
(4) What outcomes are related to feedback seeking?

Feedback Sources

The first question addressed the types of feedback sources that participants made use of to seek information about their performance during MD. Previous research (Van Dyne, 1992) has established three categories of sources of information for feedback seeking in work situations: constituencies (e.g., supervisors, coworkers, customers, subordinates), systems (e.g., tasks, work systems, job aids), and the self (one's own thoughts and feelings). The relative importance of each source to the information seeker, however, has not been clearly established. Greller's (1975) seminal study on feedback sources, for instance, showed that employees rated their supervisors as the most important source, while Hanser & Muchinsky (1978) found that employees rated their own assessment (self) as the most important feedback source. Related to assessing one's performance, self-generated information has been shown to be most frequently accessed (Greller & Herold, 1975). In fact, feedback research has shown with fair consistency, that psychologically closer sources of information, such as the self, the task, and one's peers, are most often accessed when seeking feedback, are seen as most reliable, and as easiest to obtain (Van Dyne, 1992). Extending this research to MD settings, it is important to understand what sources of feedback participants perceive as important and useful.

Individual Antecedents

The second research question centers around the individual-level antecedents of feedback seeking behavior. While many individual level variables have been proposed as potential antecedents, only a few have been substantiated in empirical studies. These are an individual's tolerance for ambiguity, which is negatively correlated

with feedback seeking, an individual's goal orientation (Madzar, 1995), and an individual's organization-based self-esteem (OBSE) which is positively correlated. Because OBSE is dependent on performance within an actual work context and might not apply in a learning setting such as MD, it was excluded from this study. Only tolerance for ambiguity and goal orientation were investigated as potential antecedents.

The literature reports negative relationships between feedback seeking and tolerance for ambiguity: individuals with a lower tolerance to ambiguity engage in more feedback seeking behavior to gain certainty about their performance (Ashford & Cummings, 1985).

Instructor-related Antecedents

Previous research (Madzar, 1995) had found that employees seek information more often from supervisors who are acting as role models, who pay individual attention to each employee, and who challenge employees to think critically. This leader behavior, known as charismatic or transformational leadership in the leadership literature, has been proposed to also apply to education and training situations (Walumbwa & Kuchinke 1999).

Reactions to Training

The fourth research question addresses the question of the effects of feedback seeking on reactions to training. Reaction measures are limited in gauging the value of training, but are valuable in this context where the primary intent was to measure the outcomes of feedback seeking behavior and not to evaluate the effectiveness of training. Assessing student reaction is "still considered appropriate for evaluating the effectiveness of training... by researchers and practitioners who have been pressed to determine the value of training activities" (Christoph, Schoenfeldt, & Tansky, 1998, p. 27). Feedback, in general, has been positively correlated with reactions to training, regardless of source. Further, information generated by the self, the task, or by via computer, have been perceived as more reliable, more trusted, and more useful than information generated by peers or supervisors (e.g.: Ang, Cummings, Straub, & Early, 1993).

Methodology

The population in this study consisted of participants and instructors of MD training programs in a U.K. Government Agency (Agency) of about 3,000 employees. The Agency had an HRD unit of 16 employees responsible for development and training services as well as internal consulting. MD constituted a key responsibility of the HRD unit because of the need to develop and retain managerial talent and to ensure a pool of qualified employees for internal promotion and succession. The focus of this study was a series of five -day training courses for employees who had been identified by their supervisors as potential future leaders. Courses were offered on average once per month and attended by 10 - 15 participants from various parts of the Agency. The courses followed a highly standardized curriculum and delivery process to ensure consistency of learning across courses and were delivered by teams of two instructors who belonged to the HRD unit. The curriculum focused on organizational issues, such as the overall strategic direction of the Agency and strategic planning and strategy implementation, and on organization behavior issues such as motivation, team building, communication, and learning styles. The courses were primarily instructor- and theory centered, but also included some role-plays, case studies, and action planning. Prior to a course, participants met with their supervisor and developed a performance contract that specified the particular performance issues on which to focus during the training. There was also in place a follow-up process designed to ensure the transfer of learning to the workplace.

The researcher observed several courses prior to the study and collaborated with HRD management and training personnel on its design. Survey data were collected from nine consecutively held courses over a seven-month period in 1998. . Five courses were taught in a residential mode and held at a seaside resort in the South of England. Four courses were taught in a non-residential mode where participants attended during working hours and then left for home. A total of 98 participants and 9 instructors completed the surveys, resulting in a response rate of over 95% for participants and 100% for instructors. MD participants were on average 34 years of age, had 11 years of professional experience, and were relatively new to their current position. Instructors, on average, were older and had lower levels of formal education but longer professional and job-related experience.

Results

Table 1 shows descriptive statistics, reliability indices, and zero-order correlations among the variables. All scales showed sufficient (Nunally, 1967) reliability. The mean scores for the four feedback sources suggested that the 98

participants in 9 MD courses did engage in feedback seeking to substantial degree (the scale anchors were 1: none, 5: lots) and that they sought feedback from a variety of sources. They did not, however, seek feedback from all sources equally. Instructors acted as the primary source of information on how well participants were meeting the learning goals of the course, followed by their peers and the course activities, and their own thoughts and feelings. This finding is in contrast to previous research where the self was the primary source of feedback.

Table 1
Course Participants' Descriptive Statistics and Zero-Order Correlations < .05 (N=98)

Variable	Mean (s.d.)	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Feedback Sources															
1.Instructors	3.60(1.05)	(.79)													
2 Peers	3.46 (.98)	.86	(.79)												
3 Course Activities	3.44 (1.07)	.76	.79	.78	(.83)										
4. Self	3.21 (.77)	.53	.61	.58	.55	(.68)									
Instructor Behavior															
5. Charismatic	4.21 (.48)	.45	.43	.38	.36	.41	(.74)								
6. Inspiring	4.44 (.50)					.31	.76	(.68)							
7. Stimulating	4.17 (.67)						.79	.75	(.72)						
8. Considerate	4.11 (.83)	.42	.34	.29			.72	.68	.75	(.76)					
Individual Variable															
9. Learning Orientation	2.89 (.61)				.24						(.70)				
9. Performance Orient.	2.09 (1.2)										(.69)				
10.Tolerance Ambig.	2.3 (.69)										-.38	(.72)			
Reaction Measures															
11. Intent to Transfer	2.14 (.55)					-.30				-.30			(.83)		
12. Motivation to Learn	2.15 (.59)	-.28	-.25	-.24		-.33					.20		.56	(.86)	
13. Reaction to Training	1.56 (.50)	-.47	-.46	-.30	-.32	-.37	.33			-.22			.57	.56	(.97)

Chronbach alpha values listed in (brackets)

Interestingly, however, the delivery mode of the course impacted the feedback sought from peers and course activities. While in both residential and non-residential courses the instructor ranked highest as a source of feedback and the Self lowest, residential course participants sought feedback from their peers and the course activities to a much higher degree than non-residential course participants. Table 2 also shows that, on average, course participants were learning oriented, that is they valued improving their abilities. They had a relatively low tolerance for ambiguity and equally low levels of intention to transfer the learning, motivation to learn, and a surprisingly low level of overall satisfaction with the training.

The four supraordinate variables were correlated highly (for feedback sources and instructor behavior) and moderately (for reaction measures and individual variables) correlated within each other. Among the variables, there were much fewer and lower correlations, with exception of moderate correlations between charismatic and, to a lesser degree, considerate instructor behaviors and feedback sources. This suggests that the more participants perceive their instructors as charismatic and considerate of their needs, the more they will seek feedback from all sources, and vice versa. An interesting finding was the negative correlations between reaction measures and feedback sources. The negative association between feedback seeking from the instructor and the other course

participants and the positive reaction training, in particular, requires careful interpretation. It suggests that participants who seek more feedback from their instructors and peers tended to be less satisfied overall with the training and vice versa. The study did not address the issue of the quality of feedback participants were able to obtain. It is possible that students were looking for information that would help them make the course more relevant but never succeeded and were therefore dissatisfied with it. This question requires closer attention in a follow-up study.

The study also addressed specific aspects of feedback seeking: the amount of information sought from each source, the frequency with which they sought feedback, and its usefulness.

Instructors, too, expected feedback-seeking behavior by the course participants to occur, that they recognized different sources of feedback, and that their estimation of the relative importance of the sources varied. Instructors, like course participants, saw themselves as a more important source of feedback than peers, activities, or the participants' own thoughts and feelings. Instructors did, however, rated their own role as providers of feedback as higher than did participants ($p < .05$), and this difference was due to an overestimation of the frequency of feedback they provided ($p < .05$). They overestimated the frequency with which they provided feedback to MD course participants.

A series of stepwise multiple regression analyses was performed with feedback source (instructor, peers, activities, and self) and feedback characteristic (amount, frequency, and usefulness) as dependent variables.

In the regression analyses, reaction to training emerged as the most important predictor to feedback seeking from Instructors, Peers, and Activities, accounting 25%, 23%, and 11% of variance respectively. In all three cases, the regression weight was negative, suggesting that participants who are not satisfied with the training tended to seek more feedback from instructors, peers, and course activities than those who were satisfied. It should be noted that this was a post-hoc survey at a single point in time. Nevertheless, the fact that the survey was taken at the end of the 5-day training course suggests that participants had a chance to reflect back over the entire week to judge both their satisfaction and feedback-seeking behaviors. Feedback seeking emerged here as a compensatory mechanism that participants employed when the course did not meet their expectations rather than the valuable resource that the literature ascribed to the construct.

Charismatic behavior by the instructor added to participants seeking feedback from him or her.

When examining the overall frequency of feedback-seeking behaviors, professional experience emerged as a very strong predictor variable, accounting for 89% of the variance. The relationship is positive and almost perfect, suggesting that those with more professional experience also tended to seek feedback more frequently than those with less experience.

Performance orientation emerged as a strong predictor variable for motivation to learn and also for intention to transfer the learning to the workplace, accounting for 51% and 50% of variance respectively. In both cases, the beta weights are positive, indicating that the higher an individual's desire to demonstrate his or her abilities, the greater the motivation to learn and to transfer the learning after the end of the MD training course. These findings were surprising given the theoretical definition and previous research that had associated higher levels of performance orientation with a decrease in the willingness to learn because individuals are primarily focused on demonstrating their abilities rather than exploring new ways of performing.

Another surprise was the strong negative correlation between intellectual stimulation and reaction to the course. Where in previous research, leader behavior that challenges individuals assumptions and encourages them to think in new ways had been shown to contribute to satisfaction with that leader, this study showed the opposite effect. The more participants perceived the instructor to challenge beliefs and assumptions, the less satisfied they were overall with the course.

Conclusions

Feedback-seeking research is an emergent strand in the organizational behavior (OB) literature and extends traditional feedback research by proposing that employees actively pursue a number of strategies to obtain feedback about their performance. This study sought to extend this line of research into a key area of HRD, management development training. Studying feedback-seeking behavior in a series of MD courses, this study suggests a number of conclusions.

First, it appears legitimate to extend feedback-seeking research to MD settings. Feedback-seeking behavior tends to occur in MD settings, where, in contrast to the OB literature, not work performance, but learning is the goal. As in regular work settings, MD participants engage in feedback-seeking behavior, they seek substantial amounts of information about their performance, they do so frequently, and seek out a variety of sources. The instructor

emerged as the primary source of feedback in terms of amount and frequency of feedback sought and its perceived usefulness. Residential courses appeared to encourage seeking feedback from peers and course activities, presumably because there more opportunities to interact than in a non-residential mode.

Second with regard to the accuracy of their own perception, MD instructors acted much like managers and supervisors in other studies, that is, they overestimated their own role in providing frequent feedback. Both roles bear similarities that might account for this: managers and instructors carry responsibility for the employee and participant performance outcomes and exert control to direct them toward these outcomes. Both groups also appear to underestimate the extent to which employees/participants self-regulate.

Third, in contrast to the OB literature, feedback seeking did not appear as a valued and positive resource. The negative association between feedback and overall satisfaction seems to suggest that those who sought more feedback were also less satisfied with the course. Feedback seeking here appears as a strategy that participants applied when the course did not meet their expectations.

Fourth, few of the hypothesized relationships based on OB literature were confirmed in this study. Charismatic behavior of the instructor was positively associated with feedback seeking, but accounted for only 10% of the variance. Professional experience emerged as a strong predictor of feedback seeking, perhaps suggesting that those with more experience were more focused on attaining specific goals and sought the information they needed to monitor their goal attainment.

Clearly, the study suggests a number of areas that need further investigation. While the self-regulation framework has been part of the social science theory base since the mid-1980s, HRD research has not yet made much use of it. The profession stands to gain from an examination of the role of self-regulation. The self-regulation framework constitutes a promising alternative to the control-oriented view of organizations and instruction. In situations where performance requirements are complex, where there are no clearly defined solutions to organizational problems, and where goals are ambiguous and preferences are shifting, external control of employee behavior is of limited effectiveness. In these situations, self-regulation is likely to be a valuable and powerful resource. Feedback seeking, as a part of the self-regulation framework can stand to receive increased attention from the HRD profession as an effective strategy in complex performance and learning situations.

This study is among the very few that investigated the role of feedback seeking in MD settings. Replications and extensions of this line of investigation are required to build a reliable knowledge base on this topic. Among the more imminent research needs are: replication of this study with other types of training and in different organizations; replication with training that is more student-centered and perhaps might provide more opportunity to seek feedback from sources other than the instructor; feedback seeking in applied problem-solving situations, such as experiential learning and action learning set with complex task without clear answers and solutions.

Finally, it is important to recognize the limitations of this study. Among them is the post-hoc design with its inherent difficulty in identifying relevant variables and attributing causal relationships. Second, there is the likelihood of single-method bias associated with using only one method of collecting information. Lastly, the sample was comparatively small and the results might reflect idiosyncrasies of the particular organization that do not generalize to other organizations.

Several implications for practice emerge from this study. First, the training of trainers should include the concepts of self-regulation and feedback seeking among course participants. Instructors who overestimate their own importance in providing feedback might fail to recognize the role of feedback from other sources. Second, participants of MD programs should be prepared to be alert to feedback from sources other than the instructor, especially their own thoughts and feelings. Upon completion of the training when participants are required to act in complex and novel situations, like leadership situations, the Self will oftentimes be the sole guidepost for assessing whether a particular course of action is appropriate or not.

References

Ang, S., Cummings, L. L., Straub, D. W., & Early, P. C. (1993). The effects of information technology and the perceived mood of the feedback giver on feedback seeking. Information Systems Research, 4(3), 240-261.

Ashford, S. J., & Cummings, L. L. (1983). Feedback as an individual resource: Personal strategies of creating information. Organization Behavior and Human Performance, 32, 370-398.

Ashford, S. J., & Cummings, L. L. (1985). Proactive feedback seeking: The instrumental use of the information environment. Journal of Occupational Psychology, 58, 67-79.

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Englewood Cliffs, NJ: Prentice Hall.

BEST COPY AVAILABLE

Bass, B. M., & Avolio, B. J. (1995). Multifactor leadership questionnaire: Manual leader form, rater, and scoring key for MLQ (Form 5x-Short). Redwood City, CA: Mind Garden.

Bassie, L. J., & Van Buren, M. E. (1998, January). The 1998 ASTD state of the industry report. Training and Development, 23 - 43.

Brinko, K. T. (1993). The practice of giving feedback to improve teaching: What is effective? Journal of Higher Education, 64(5), 574 - 593.

Christoph, R. T., Schoenfeldt Jr., G.A., & Tansky, J. (1998). Overcoming barriers to training utilization technology: The influence of self-efficacy factors on multimedia-based training receptiveness. Human Resource Development Quarterly, 9(1), 25-38.

Conover, D. K. (1997). Leadership development. In R. L. Craig (Ed.), The ASTD training and development handbook, (4th ed) (pp. 581-600). New York: McGraw-Hill.

de Bettignies, H.C. (1975). Management develoment: The international perspective. In B. Taylor and G. L. Lippitt (Eds.), Management development and training handbook (pp. 3 - 11). New York: McGraw-Hill.

Dweck, C.S. (1986). Motivational processes affecting learning. American Psychologist, 41, 1040-1048.

Goldstein, I. L. (1993). Training in organizations (3rd ed.) . Pacific Grove, CA: Brooks/Cole.

Greller, M. M., & Herold, D. M. (1975). Sources of feedback: A preliminary investigation. Organizational Behavior and Human Performance, 13, 244-256.

Greller, M.M. (1980). Evaluation of feedback sources as a function of role and organizational level. Journal of Applied Psychology, 65, 24-27.

Hanser, L. M. & Muchinsky, P. M. (1978). Work as an information environment. Organization Behavior and Human Performance, 21, 47-60.

Holton, E.F.III;, Bates, R., Seyler, D. & Carvalho, M. (1997). Towards a construct validation of a transfer climate instrument. Human Resource Develoment Quarterly, 8(2), 95-113.

Ilgen, D. R., Fisher, C.D., & Taylor, M.S. (1979). Consequences of individual feedback on behavior in organizations. Journal of Applied Psychology, 64, 349-371.

Institute of Personnel and Development, (1999, April). Training and development in Britain 1999. IDP survey report No. 3. London: author.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychological Bulletin, 199(2), 254-284.

Kowitz, G. T., & Smith, J. C. (1985). The dynamics of successful feedback. Performance and Instruction, 24(8), 4 - 6.

Kuchinke, K. P., Brown, J. M., Anderson, H., & Hobson, J. (1998). Assessing training needs of manufacturing employees in rural Minnesota: A model and results. Journal of Industrial Teacher Education, 36(1), 79-98.

Madzar, S. (1995). Feedback seeking behavior: A review of the literature and implications for HRD. Human Resource Development Quarterly, 6(4), 337-349.

Nunnally, J. C. (1967). Psychometric theory. New York: McGraw-Hill.

Parsons, M. B., & Reid, D. H., (1995). Training residential supervisors to provide feedback for maintaining staff teaching skills with people who have several disabilities. Journal of Applied Behavior Analysis, 28, 317-322.

Schoenfeldt, E. L. (1996). Goal setting and feedback as a posttraining strategy to increase the transfer of training. Perceptual and Motor Skills, 83, 176-178.

Taylor, M. S., Fisher, C.D., & Ilgen, D. R. (1984). Individuals' reactions to performance feedback in organizations: A control theory perspective. In K. M. Rowland & G.R. Ferris (Eds.), Research in Personnel and Human Resources Management, Vol. 2 (pp. 81-124). Greenwich, CT: JAI Press.

Van de Ven, A. H., & Grazman, D. N. (1995). Technological innovation, learning, and leadership. Discussion paper, Strategic Management Research Center, University of Minnesota, Minneapolis.

Van Dyne, L. (1992). Feedback sources and types in organizational settings: typology and process model. Unpublished manuscript, Department of Strategic Management and Organization, Carlson School of Management, University of Minnesota: Minneapolis.

Vandewalle, D. (1995). A goal orientation model of feedback seeking and performance. Unpublished manuscript, Department of Strategic Management and Organization, Carlson School of Management, University of Minnesota: Minneapolis.

Viau, R., & Clark, R. (1987, April). Guidelines from research: Feedback during training. Performance and Instruction, 6 - 7.

Walter, G. M. (1996). Corporate practices in management development. Report No. 1158-96-RR. New York: The Conference Board.

Walumbwa, O. F. & Kuchinke, K. P. (1999, March). HRD faculty as leaders: The application of the full range leadership theory to graduate level HRD instruction. Paper presented at the Academy of Human Resource Development conference, Fairfax, VA.

Weinstein, (1994). Design and development of an assessment of readiness for training: The START [Strategic Assessment of Readiness for Training]. Paper presented at the Academy of Human Resource Development conference, St. Antonio, TX.

Figure 1: Theoretical Framework

A Five Phase Framework for Designing a Successful Multirater Feedback System

Allan H. Church
W. Warner Burke Associates Inc.

Janine Waclawski
W. Warner Burke Associates Inc.

Given the increasing popularity and application of multirater feedback methodology for both developmental and decision-making purposes in organizations there is a need in the literature for more attention to the factors involved in creating a successful feedback process. The following paper describes a five-phase framework for designing such a MRF system based on the classic OD consulting skills model and many years of practitioner experience with large-scale feedback-based applications in Fortune 100 organizations.

Keywords: Multirater Feedback, Behavior Ratings, Consulting Framework

Multirater feedback methodology is one of the most rapidly advancing areas in HRD, OD and I/O practice and research today. The 1990s has been heralded by some as the age of information and thus it should come as no surprise that perceptions, opinions and ratings gathered from multiple sources at work such as managers, direct reports, peers, team members, colleagues, supervisors (straight and/or dotted-line), and clients or customers has emerged as one of the most popular and prevalent developmental and assessment tools in organizations (Bracken, 1994, 1996; Borman, 1997; Church, 1995; Church & Waclawski, 1998a; London, 1997; McLean, 1997; Tornow, 1993; Waldman, Atwater & Antonioni, 1998; Yammarino & Atwater, 1997).

Also known as multisource, 360-degree and full circle feedback, the uses of this type of information appear to be almost as endless as the number of different names or the number of sources from which data can be collected. A basic review of the literature, for example, yields feedback applications related to such areas as performance appraisal via upward or full 360-methods (e.g., Antonioni, 1996; Edwards & Ewen, 1996), individual and organizational development and change efforts (e.g., Burke, Richley & DeAngelis, 1985; Church, Javitch & Burke, 1995; Church & Waclawski., 1999; Church, Waclawski, & Burke, 2000; London & Beatty, 1993; Nowack, 1992), executive development and coaching methods (Dalton, 1996; Goodstone & Diamante, 1998; Hazucha, Hezlett & Schneider, 1993; Waclawski & Church, 1999), and even large-scale cultural assessments and trend analyses (e.g., Church, 1999a; London, 1997).

Despite, or perhaps even because of, the widespread adoption of this methodology in various settings and by all kinds of individuals (including managers and other types of consultants), many practitioners have begun to raise important questions about the design and implementation of these systems. While feedback applications are indeed abundant, given the reported failure rates of organizational change efforts in general and related management consulting fads such as TQM and reengineering efforts (e.g., Kotter, 1995; Schaffer & Thomson, 1992; Spector & Beer, 1994; Trahant & Burke, 1996), questions are now being raised about what works and what does not with respect to multirater feedback systems as well. Given the millions of dollars that are being invested in the design of such technology, often created specifically for a given organizational context, there is need to move beyond the hype associated with 360-degree feedback and focus instead on developing a successful system that will indeed promote behavior change and help improve individual and organizational performance. Moreover, meta-analytic research on the impact of feedback in general (Kluger & DeNisi, 1996) has shown that simply handing people some results in and of itself does not result in positive change, and can in fact cause decreased effectiveness over time.

What is critical, however, is the why, how, what and when involved in the process. That is to say, there is a combination of factors that contribute to the successful utilization of multirater feedback for positive individual and organizational change. Some of these elements include gaining support and commitment from senior leadership, creating meaningful linkages with corporate strategy and objectives, establishing feedback as an ongoing process

Copyright © 2000, Allan H. Church & Janine Waclawski

not just a single event, the method of delivery (e.g., vis-à-vis coaching vs. the non-effective "desk-drop" used by some), and ensuring that action takes place as a result of the personal findings.

While a number of practitioners have suggested critical factors, guidelines, and potential external variables that might influence the success of a multirater effort (e.g., Antonioni, 1995; Church & Bracken, 1997; Church & Waclawski, 1998a; McLean, 1997; Waldman et al., 1998; Wimer & Nowack, 1998), what is still needed in the literature at this point is clear, concise, practitioner framework for internal and external HRD, OD and I/O professionals involved in designing a developmental feedback system. Only by making a concerted effort to attend to the design process itself can we move toward a better understanding of why some multirater feedback efforts fail and some are highly successful. Such a practitioner framework would need to delineate the important factors in each aspect of the process from initial development, through feedback delivery, to evaluation. Although models of this type exist for other types of assessment methodologies such as organizational opinion surveys (e.g., Church & Waclawski, 1998b; Nadler, 1977), there is a need for a similar comprehensive approach to feedback applications.

The purpose of the following paper, then, is to introduce such an applied framework that will be helpful to both practitioners and researchers alike based on a combination of prior theory, research, and consulting practice regarding the intersection of the areas of organization change and development and multirater feedback. Following a brief overview of the core assumptions of multirater feedback methodologies inherent in this type of approach, a five-phase practitioner framework will be described in detail. Although the model is based on a combination of existing literature and 15 years of combined consulting experience with such systems, and therefore is directly relevant to HRD and OD practice, areas for research and further development will also be highlighted. Finally, limitations of the model will also be discussed.

Core Assumptions of MRF

In general, the process of collecting and utilizing data from multiple sources for individual development, performance improvement, and organizational change is based on several key assumptions. One set of these assumptions reflects the more methodological aspects of the assessment process, and the second concerns the psychological and cognitive processes involved. Each of these is described briefly below.

First, there is the assumption of *behavioral consistency*. Based on measurement theory (Landy & Farr, 1980; McLean, 1997) the premise is relatively simple. In order to provide an individual leader, manager or executive with feedback, the behavior of that individual must be observable and subsequently quantifiable using some form of scale or index. Fundamentally, then, there is an inherent belief here that individuals do in fact behave in certain, consistent, and observable ways with co-workers, clients and/or other organizational members, thus making ratings possible. Hence, multiple ratings from the same (within) observer source will ultimately increase the validity and reliability of the information obtained. At the same time, however, it is also assumed that individual behavior differs enough based on the nature of the situation, role sets and other individuals involved in the process so that multiple ratings from different perspectives are indeed worth collecting. From this perspective, ratings from different (between) sources such as those of direct reports, peers or customers will result in a greater range of opportunities for observation and subsequently behaviors being assessed for comparison purposes (Harris & Schaubroeck, 1988; Landy & Farr, 1980). These two elements of behavioral consistency, of course, lead to the other primary methodological assumption of *observer consistency*. In short, observer consistency means that observers (or co-workers, raters, etc.) have the appropriate skill, ability, and data collection methodology to accurately assess, index and record the behavior of the focal individual. Clearly, without these two assessment related assumptions, the process of collecting behavior data is relatively pointless. Despite this, many MRF systems are developed with little to no attention to who the appropriate raters might be or whether or not they are in a position to accurately assess the behaviors of the focal individual.

The second major set of assumptions in MRF concerns the psychological and cognitive processes involved. First, there is a need for *positive rater motivation*. On the part of the observers, this is reflected in a relatively straightforward assumption— i.e., that raters are appropriately motivated to provide accurate ratings. Similarly, the motivation of the focal individual involved is also assumed to be positive in nature— i.e., he or she is interested in receiving accurate feedback with the intent to use it for personal development and/or performance improvement. In practice, of course, motivation is always an issue, particularly when MSF is mandated by senior management without appropriate forethought and linkages to existing organizational systems and initiatives (Harris, 1994). A second major cognitive psychological assumption in MSF is the notion that *feedback enhances self-awareness*. Based on a combination of clinical and social psychological theories (e.g., Argyris, 1970; Bion, 1959; Lewin,

1958), the conceptual chain of events begins with an individual assessing his or her own behavior (thus the need for a self-rating, which not all applications of MSF employ) prior to receiving feedback. This sets the cognitive stage via self-perceptions that, when compared with some collective of others' observations, initially results in cognitive dissonance (Festinger, 1957) and ultimately greater clarity regarding (a) how others' experience this individual's behavioral strengths and weaknesses in the workplace, (b) how these compare with this individual's own perceptions on the same managerial or leadership facets, and typically (c) how both sets of data compare with some larger comparison group or norm such as those collected from other workgroups, departments, functions, or even other organizations. Following initial resistance and rejection, the assumption here is that this process of greater clarity can, in turn, given the appropriate environment and tools, result in energy and impetus for change (Lewin,1958; London, 1997) as well as the potential for enhanced feedback seeking activity (Ashford & Tsui,1991) in the future. This process has been described in detail elsewhere (Church & Waclawski, 1999).

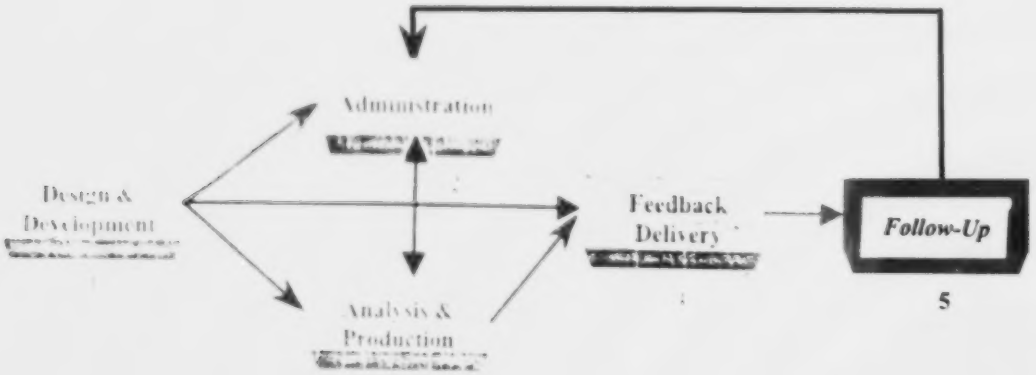
Clearly, the above assumptions require that a number of conditions and skills exist for an effective MRF initiative to work as intended. Next, we will describe a five phase framework for designing MRF systems that addresses these factors.

Five Phase Framework of Multirater Feedback

The five phase framework of multirater feedback is adapted in part from the classic OD model of organizational consulting (Kolb & Frohman, 1970). Based on the action research perspective that highlights the importance of data in any change process (Lewin, 1946; 1958), it consists of seven clear phases; phases, not steps, since there is overlap throughout the process. The seven phases in the classic consulting model are, *entry, contracting, data collection, data analysis, data feedback, intervention, and evaluation* . These phases represent the complete process from beginning to end in any given consulting relationship (e.g. OD, I/O, HRD or whatever).

The five phase framework for MRF is similar to the OD model in that (a) it is also phased-based so there is an open flow throughout the process, and (b) it follows a natural progression from the initial ground work through the data collection and delivery and requires follow-up. Although relatively intuitive and straightforward in appearance (see Figure 1), the issues inherent in each phase are important and often overlooked when practitioners and others design MRF systems. Each of the five phases is described briefly below along with some of the key issues and challenges involved.

Figure 1. Five Phases of MRF Implementation



Design and Development

The first step in any significant organizational undertaking is to ensure that you have a firm understanding of the context involved. The design and development of a new feedback process (whether the tools and methods are in fact to be developed from the ground-up or adapted from some existing source) is no exception. Thus, before the internal or external feedback practitioner can begin to design and develop the actual assessment instrument, it is essential that two critical concerns be addressed: (1) He or she must understand the organizational environment

BEST COPY AVAILABLE

(i.e., the systemic and cultural contents) in which the MRF system is being created and delivered, and (2) he or she must have a firm grasp of the client organization's needs vis-à-vis the development and use of the feedback system. These two factors more than any other determine the *relevance* and ultimate *impact* of the feedback, as no MRF system is created or used in a vacuum. Understanding the key components of the organization (including the mission, strategy, culture, and existing systems), as well as the specific needs of the client are essential to developing a meaningful and successful MRF system.

Prior experience here is one importance factor. For example, if employees have had little exposure to receiving any kind of formal individual feedback, such as might be the case in a professional services firm (Church et al., 1995), the nature of the communications and leadership efforts needed to model and support the feedback process will be different than in an organization that is very familiar with such methods. Establishing linkages to existing initiatives and performance objectives is also critical. For example, a feedback system that is directly linked to the organization's strategic objectives and has been designed at the request of the client to help move the organization and its leaders forward in the accomplishment of these objectives is likely to be significantly more impactful than a feedback system derived from some standard set of general management practices.

Also important in this phase, of course, are the fundamental principles and critical is sues involved in creating a MRF instrument, such as obtaining buy-in" and support from key organizational members, creating a MRF system plan, identifying and understanding the target audience for feedback, item writing and scale construction, readability issues, and the importance of considering the type of action planning efforts to be utilized when initially constructing the instrument. The latter skills, though often overlooked by many leaders, managers and even some HR professionals with little methodological experience, can nonetheless have a significant impact on the utility and therefore the ultimate success of the results obtained. This is one of the reasons that successful MSF efforts typically require good, solid experience in item writing and scale design as well as attention to the larger more systemic factors.

Administration

This second phase consists of all the logistical and tactical elements required to collect the MRF data. Important concerns for the practitioner include how to effectively communicate the purpose and objectives of the MRF system to those involved (i.e., raters, ratees, and those in a position to reinforce the process), options for administration and data collection (e.g., on line, paper and pencil, email, voice response unit, etc.), how to coordinate with the client organization to ensure a smooth and seamless administration of the feedback instrument, proven techniques for following up with participants and respondents to ensure the highest possible response rates, best practices regarding the timing of the administration and its impact on response rates, the importance of reinforcing and protecting confidentiality during the entire feedback process, and suggestions for working with clients who have not had previous experience in administering MRF systems, which, despite the popularity of MRF, still represents the majority of organizational settings.

In many ways this phase also contributes significantly to the success or failure of the MRF effort. For example, if the participant identification codes are not correct for each focal manager and his or her respondents, the integrity and validity of the MRF results obtained will be void. Without a well thought out and implemented plan for administering the MRF paperwork (or tracking online responses) the entire MRF process will be invalidated before it even begins. Moreover, making sure that participants are kept informed about the response rates of their constituents is also essential to the process. Specifically, good response rates are necessary to ensure the most complete and therefore accurate ratings possible. Without sufficient responses from key constituents (for example, supervisors are often slower than other groups to respond), the perceived quality and usefulness of the feedback is at risk. Although some research (Church, Rogelberg & Waclawski, 1998) has suggested that there may only be a very minor negative relationship between prior performance and response rates in MRF efforts, these are importance concerns nonetheless from the participant and coaching perspectives.

Finally, the importance of confidentiality in the MRF process can not be overstated. Participants' and especially their respondents' belief in the anonymity of their responses is the cornerstone of a good MRF system. From the participant or focal manager perspective, the confidentiality of the final results themselves is critical (i.e., no one else within the participant's organization will have access to his or her results without his or her authorization). From the respondent perspective, a firm belief that their individual ratings will in no way be identified or revealed to the focal manager receiving feedback is paramount to ensure open, candid and sometimes

less than favorable ratings of the focal manager's behavior. These are issues that must be managed continually throughout the administration process and well into the later stages as well.

Analysis and Production

This phase is focused on the actual generation or creation of the feedback report itself. There are many issues for the HRD, OD and I/O practitioner to consider and address here, the most important of which are those related to the look and feel of the report. These elements are critical as they directly impact the client's (or feedback recipient's) reactions to and understanding of the MRF data. For example, in our experience feedback reports that contain both numeric and graphic data displays are far superior to those that contain either one or the other. We have found this to be the case because people have different orientations to reading, interpreting and working with data. Quite simply some prefer numbers ands others prefer charts. By providing both types of displays for the participant, maximal usefulness is guaranteed. The same argument can be made for providing various types of text interpretation (and/or the additional use of write-in comments either in "cleaned" verbatim or content coded form) in addition to more traditional use of mean score results in various formats.

Another critical element in report analysis and production is the nature of the information being presented. In general we advise the use of average rating scores for each perspective (to ensure confidentiality) in conjunction with some kind of indication of the range of responses (to provide specificity). Typically, the range is better reflected via a form of highest and lowest indicator rather than as a full distribution of individual scores which tends to allow participant's to speculate too much regarding "who provided that particularly low rating." The key is to provide the participant with the most useful (in other words specific) data you can without jeopardizing the confidentiality of the process. Often this is a delicate balance.

Finally, the "user friendliness" of the feedback report itself cannot be overlooked. Specifically, it is essential that no matter what type of data is being displayed in whatever form it must be intelligible to the participant or focal manager. At the most basic level, this means clearly labeling each item number, presenting the questions as they were asked, and linking these to graphic or visual data clearly on each participant's report. Moreover each should display the number of respondents who participated in the feedback process (by each constituency group— e.g., direct reports, peers, clients, supervisors, etc.), group normative data (for the larger function, unit, region or even organization as a whole), and of course the name of the focal manager being rated (we strongly suggest this be included on every single page). Attention to details such as these is critical in ensuring increased readability and therefore increased usefulness of the MRF report.

In short, no matter how well constructed and administered the MRF instrument, if the final report is not accurate, clear and easy to read and work with, the MRF initiative will fail to be a successful one. Some additional critical factors that will also be discussed include the importance of assessing the client's sophistication and experience in working with data and feedback reports prior to delivery, the speed and quantity of production methods and systems (i.e., how many people will need reports and in what time frame), and selecting the best and most user friendly formats for data presentation.

Feedback Delivery

This phase consists of the actual delivery or distribution of the MRF report to feedback recipients throughout the organization. There are a variety of modes of delivery available to the feedback practitioner that vary on several key dimensions: (1) the cost of delivery, (2) the desired level of integration with other organizational initiatives, and (3) the degree of assistance and support to be provided to the focal individual (feedback recipient) to help him or her understand, synthesize, and ultimately take action from the results. Each of the possible modes of delivery— e.g., from a multi-day residential coaching based program with biannual follow-up coaching efforts to the infamously ineffectual "desk-drop"--and their associated advantages and disadvantages will be discussed.

Until very recently executive development has been almost exclusively conducted as a group process. Historically, executives and managers have been developed through training sessions resembling classroom learning. With the enormous growth (over the past decade) in the use of individual assessment in the workplace, the use of multirater feedback as part of a company's formal executive development process has become quite commonplace. Nevertheless, even though the delivery of this type of feedback is usually done within the context of a group training session (such as multi-day residential development program), the focus of the feedback is invariably on the perceptions of the *individual* executive--specifically, how he or she is seen by direct reports,

peers, supervisors, and even clients and/or customers. So while the feedback is delivered in a group setting, the focus is typically on the individual. In fact, in the present five phase model of MRF, coaching is integral to the process of learning and ultimately changing.

However, despite the popularity of such large group oriented developmental programs, many managers simply do not have the time they once had to invest in attending training programs that last several days. Thus, in some situations, individual coaching with MRF as the primary input is often used independently in lieu of large group training and development efforts. Although this approach can provide added benefits in terms of greater individualized attention to the focal manager, downsides can include missing out on group learning and networking experiences, missed opportunities to share thoughts, feelings and insights based on feedback, the absence of group support, and discussions about the impact of the organization's culture on giving and receiving feedback. Of course, individual coaching done outside the context of a large group training session can also be much more costly to an organization as economies of scale are almost nonexistent given the associated travel costs coupled with many consultants' high daily rates.

The final mode of delivery, known as the "desk drop", simply consists of providing each focal manager with a copy of his or her feedback report sans coaching or any other helpful interpersonal input. Although this method is the least costly, it is also in our experience the least effective. Participants are left to their own devices to decipher and make behavioral changes based on their personalized feedback. No explanation, assistance or guidance is provided in this type of delivery. Needless to say, we do not recommend this approach as we feel it is potentially more harmful than providing no feedback at all, and more than likely represents a waste of resources .

Follow-Up

While this seemingly "final" phase is critical as it marks the completion of the MRF cycle, unfortunately, as with most change efforts, it often receives far too little attention from practitioners and their client organizations. This is disconcerting as following up to assess the impact of any MRF initiative (or any organizational improvement effort in general) is critical not only to the perceived credibility of the current MRF initiative, but also to the viability of future ones as well as the ultimate integrity of the HRD, OD or I/O function associated with it. Quite simply demonstrating impact by measuring progress and examining the link between MRF and other measures of individual and organizational performance (Church, 1995b; Waclawski, 1996) is vital to establishing the importance and credibility of the MRF or any related assessment process. To this end, techniques for assessing impact, methods for linking MRF data to other measures of performance, and the importance of creating long-term performance systems and strategies that include the use of MRF are discussed below.

As a rule, the criteria one uses to assess the success of any MRF process should be determined well in advance of the follow-up phase. In fact, these criteria should be determined before the MRF process even starts. More specifically, the MRF process should be driven by some quantifiable set of deliverables and/or outcomes such as better leadership, increased productivity, enhanced communication, or improved team climate and effectiveness etc. If these objectives are not determined in advance of the instrument construction and then built into the feedback effort itself, they cannot be achieved let alone evaluated as part of the follow-up process. Therefore, careful though and consideration must be given to the purpose and end products of any MRF effort well in advance of its administration.

In addition to clearly identifying and operationalizing outcome criteria at the beg inning of the MRF process, systems which actually *support* and *reinforce* individual behavior change must also be put into place. Coaching, as we have already discussed, is one example of such a support mechanism that reinforces individual change, especially when it is offered as an on-going process as opposed to a single coaching session. Internal organizational resources such as additional training and on the job skill development also serve to increase the likelihood of lasting changes based on MRF results. Unfortunately, all too often this critical factor is overlooked or rejected at some point during the MRF planning and budgeting process. Thus, in the end in many situations, the expense of creating such organizational support systems is often the determining factor in the decision not to implement them. In our experience, MRF efforts that are conducted without institutionalized systems to support change are likely to be more costly in the long run than those with only adequate support, as they going to be are far less effective.

Finally, there are two ways to link MRF to predetermined performance criteria. The first is to embed MRF practices or behaviors in other existing appraisal and/or measurement processes (i.e., the performance appraisal system, employee and/or external customer satisfaction surveys, etc.). More specifically, this means taking key

items from the MRF questionnaire and making them a part of other important and well utilized organizational measurement and assessment tools. By making these formal connections and linkages, the focal manager will not only receive feedback on the key practices in a developmental context (through the MRF process), but he or she will also receive pay and promotion increases (or decreases) as a result of his or her performance on these same key indicators.

The second approach to linkage is through correlational and/or regression analysis (e.g., Church, 1999b; Waclawski, 1996). This involves linking MRF feedback at the individual, group, and organizational levels to other indices (typically hard measures) of performance such as ratings of turnover, absenteeism, departmental productivity, sales, ROI, customer satisfaction, repeat business, etc. This approach capitalizes on linking existing databases of critical organizational performance measures to an MRF (or survey) instrument and is simpler in some ways to conduct as it does not require embedding MRF practices into preexisting measurement tools. Unfortunately, such an approach does represent its own constraints as well, including an inability to determine true causal relationships. In general, however, both approaches are useful in reinforcing the importance of behavioral change by demonstrating both to organizational members and its leadership that job performance is contingent not only on measures of bottom line productivity but also on behavior; specifically, how each focal manager behaves as a leader and or manager. Without creating a "hard" link between behavioral change and job performance ratings, the importance of the MRF process to job performance cannot easily or credibly be established.

Limitations and Areas for Future Research

No framework or model is perfect and the present approach to MRF is no exception. While our model may seem overly simplistic and perhaps lacking in sufficient detail, it does provide the would-be MRF practitioner and even the MRF client with a good place to start. With respect to future research, our aim is to further develop and study this model and its applications across a variety of settings and contexts. In particular, we believe that establishing the link between MRF and organizational performance will be of critical importance in the coming decade. We also believe that this link will be critical not only to MRF researchers but will in large part determine the longevity (and continued viability) of the MRF methodology itself, as it will for many different types of OD and HRD related efforts. In short, if we as a field do not successfully demonstrate the linkages between the work we do and the continued improvement of individuals and organizations, the perceived utility of our efforts in organizations will indeed wane. To a large extent this has begun to happen already with respect to other similar forms of behavioral and attitudinal assessment such as organization surveys. MRF is simply the next arena.

References

Antonioni, D. (1995). Problems associated with implementation of an effective upward appraisal feedback process: An experimental field study. Human Resource Development Quarterly, 6(2), 157-171.

Antonioni, D. (1996). Designing an effective 360-degree appraisal feedback process. Organizational Dynamics, 25(2), 24-38.

Argyris, C. (1970). Intervention theory and method. Reading, MA: Addison-Wesley.

Ashford, S. J., & Tsui, A. S. (1991). Self-regulation for managerial effectiveness: The role of active feedback seeking. Academy of Management Journal, 34, 251-280.

Bion, W.R. (1959). Experiences in groups. New York: Basic Books.

Borman, W. C. (1997). 360-dergee ratings: An analysis of assumptions and a research agenda for evaluating their validity. Human Resource Management Review, 7(3), 299-315.

Bracken, D. W. (1996). Multisource (360-degree) feedback: Surveys for individual and organizational development. In A. I. Kraut (Ed.), Organizational surveys: Tools for assessment and change (pp. 117-143). San Francisco: Jossey-Bass.

Bracken, D. W. (1994). Straight talk about multirater feedback. Training & Development, 48(9), 44-51.

Burke, W. W., Richley, E. A., & DeAngelis, L. (1985). Changing leadership and planning processes at the Lewis Research Center, National Aeronautics and Space Administration. Human Resource Management, 24, ... 90.

Church, A. H., (1999a). Large scale applications of small scale data. Performance in Practice, (Spring), 7-8.

Church, A. H., (1999b). Do higher performing managers actually *receive* better ratings? A validation of multirater assessment methodology. K. P. Kuchinke (Ed.), Academy of Human Resource Development 1999 Conference Proceedings, Vol. 2., 791-798, Baton Rouge, LA: AHRD.

Church, A. H. (1995a). First-rate multirater feedback. Training & Development, 49(8), 42-43.

Church, A. H. (1995b). Linking leadership behaviors to service performance: Do managers make a difference? Managing Service Quality, 5(6), 26-31.

Church, A. H., & Bracken, D. W. (1997). Advancing the state of the art of 360-degree feedback: Special issue editors' comments on the research and practice of multirater assessment methods. Group & Organization Management, 22(2), 149-161.

Church, A. H., & Javitch, M., & Burke W. W. (1995). Enhancing professional service quality: Feedback is the way to go. Managing Service Quality, 5(3), 29-33.

Church, A. H., Rogelberg, S. G., & Waclawski, J. (1998). Do high performing managers get more or less multirater feedback? Paper presented at the 13th Annual Meeting of the Society for Industrial and Organizational Psychological (SIOP), April 25, Dallas, Texas.

Church, A. H., & Waclawski, J. (1999). Influence behaviors and managerial effectiveness in lateral relations. Human Resource Development Quarterly, 10(1), 3-34.

Church, A. H., & Waclawski, J. (1998a). Making multirater feedback systems work. Quality Progress, 31(4), 81-89.

Church, A. H., & Waclawski, J. (1998b). Designing and using organizational surveys. Aldershot, England: Gower.

Church, A. H., Waclawski, J., & Burke, W. W. (2000). Multisource Feedback for Organization Development and Change. In D. W. Bracken, C. W. Timmreck, & A. H. Church (Eds.). The Handbook of Multisource Feedback, in press, San Francisco, CA: Jossey-Bass

Dalton, M. (1996). Multirater feedback and conditions for change. Consulting Psychology Journal: Practice and Research, 48(1), 12-16.

Edwards, M. R., & Ewen, A. J. (1996). 360-degree feedback. The powerful new model for employee assessment and performance improvement. New York: AMACOM.

Festinger, L. (1957). A theory of cognitive dissonance. Evanston, IL: Row-Peterson.

Goodstone, M. S., & Diamante, T. (1998). Organizational use of therapeutic change: Strengthening multisource feedback systems through interdisciplinary coaching. Consulting Psychology Journal: Practice and Research, 50(3), 152-163.

Harris, M. M., (1994). Rater motivation in the performance appraisal context: A theoretical framework. Journal of Management, 20(4), 737-756.

Harris, M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. Personnel Psychology, 41, 43-61.

Hazucha, J. F., Hezlett, S. A., & Schneider, R. J. (1993). The impact of 360-degree feedback on management skills development. Human Resource Management, 32(2-3), 325-351.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis and a preliminary feedback intervention theory. Psychological Bulletin, 119(2), 254-284.

Kotter, J. P. (1995). Leading change: Why transformation efforts fail. Harvard Business Review. 73(2), 59-67.

Landy, F. & Farr, J. (1980). Performance rating. Psychological Bulletin, 87, 72-107.

Lewin, K. (1958). Group decision and social change. In E. E. Maccoby, T. M. Newcomb, and E. L. Hartley (Eds.), Readings in Social Psychology (pp. 197-211), New York: Holt, Rinehart, and Winston.

Lewin, K. (1946). Action research and minority problems. Journal of Social Issues, 2, 34-46.

London, M. (1997) Job Feedback: Giving, seeking and using feedback for performance improvement. Mahwah, NJ: Lawrence Erlbaum Associates.

London, M., & Beatty, R. W. (1993). 360-degree feedback as a competitive advantage. Human Resource Management, 32(2&3), 353-372.

McLean, G. N. (1997). Multirater 360 feedback. In L. J. Bassi & D. Russ-Eft (Eds.) What works: Assessment, development, and measurement (pp. 87-108). Alexandria, VA: American Society for Training and Development.

Nadler, D. A. (1977). Feedback and organization development: Using data-based methods. Reading, MA.: Addison-Wesley.

Schaffer, R. H., & Thomson, H. A. (1992). Successful change programs begin with results. Harvard Business Review, 70(1), 80-89.

Spector, B., & Beer, M. (1994). Beyond TQM programmes. Journal of Organizational Change Management, 7(2), 63-70.

Tornow, W. W. (1993). Editor's note: Introduction to special issue of 360-degree feedback. Human Resource Management, 32(2-3), 211-219.

Trahant, B., & Burke, W. W. (1996). Creating a change reaction: How understanding organizational dynamics can ease reengineering. National Productivity Review, Autumn, 37-46.

Waclawski, J. (1996). Using organizational survey results to improve organizational performance. Managing Service Quality, 6(4), 53-56.

Waclawski, J. & Church, A. H., (1999). Four easy steps to performance coaching. Performance in Practice, (Summer), 4-5.

Waldman, D. A., Atwater, L.E., & Antonioni, D. (1998) Has 360 degree feedback gone amok? Academy of Management Executive, 12(20) 86-94.

Wimer, S., & Nowack, K. M., (1998). 13 common mistakes using 360-degree feedback. Training & Development, 52(5), 69-78.

Yammarino, F. J., & Atwater, L. E. (1997). Do managers see themselves as others see them? Implications of self-other rating agreement for human resources management. Organizational Dynamics, 25(4), 35-44.

An Evaluation of the Quality of 360-degree Assessment Instruments

Froukje Jellema
Adrie Visscher
Martin Mulder
University of Twente, The Netherlands

Recently 360-degree assessment has become popular. It can be used as well for evaluating employee performance as for the evaluation of training transfer. The instruments used for this kind of assessment, where managers are evaluated by co-workers such as their supervisor, peers, subordinates and clients, should meet certain standards (e.g. psychometric and personnel evaluation standards). A checklist of standards will be presented on the basis of which 360-degree instruments are examined. In this study four 360-degree assessment instruments are examined. The outcomes of the study give insight into the qualities of a number of 360-degree instruments.

Keywords: 360-Degree Assessment, Multirater Assessment, Training Evaluation

Introduction and Problem Statement

In a multirater assessment, job-performance of an employee is evaluated by one or more co-workers, such as supervisors, subordinates, peers, customers, and suppliers. If all of these sources are being used, and the employee also evaluates his or her own job performance, this is called a 360-degree (full circle) assessment. The assumption is that with each additional rater source, the confidence that the reported results are an accurate reflection of what is happening is increased (Robinson and Robinson, 1989).

360-degree assessment can be used for several purposes. First, it can be used as an instrument for personnel development, for example for analyzing the strengths and weaknesses of an employee. It can also, as an element of a training or development program, provide a trainee with a clear picture of his or her performance and training priorities. Furthermore, 360-degree assessments can be instruments in formal performance appraisal, for example as a basis for making salary or promotion decisions. Finally, 360-degree assessments could be used in training evaluation. In that case, the trainee's co-workers provide feedback, both before and after the training, and/or are asked directly for perceived changes in their colleague's job-performance, as a result of the training he or she has undergone.

The 360-degree assessment usually concerns a questionnaire, by means of which raters give their opinion about the ratee's job-performance. This questionnaire generally consists of several competencies that are considered relevant for the ratee's job-performance. For example, an instrument may focus on five groups of competencies, i.e. leadership, communication, management, decision making, and personal behavior. Each of these groups is measured by more specific behavioral characteristics. For instance, 'communication' may be measured by 'listening skills', 'written communication', 'oral communication' and 'presentation skills'. Each category is usually measured by means of several items.

Ten years ago, only twenty to thirty instruments were on the US market, now over one hundred are available (Lepsinger and Lucia, 1997). The quality of these instruments should be ascertained since it will not matter what the data collection process reveals about an individual, if the instrument lacks validity, reliability and applicability for the organization (Church, 1999). Some instrument evaluation has already been carried out (Morical, 1999; Van Velsor and Leslie, 1991). Several studies have focused on their reliability, (Nijhof and Jager, 1995) and validity (Church, 1999), and on the effects of 360-degree instruments for evaluating management development (Rosti and Shipper, 1998; McLean et al., 1995; Hazucha et al., 1993). However, most of these studies focus on US-instruments.

The increasing use of 360-degree assessments in Dutch organizations has resulted in the development of a considerable number of Dutch instruments. This popularity, however, has not been supported by research on their quality. This study is therefore meant to reveal more about the extent to which Dutch instruments meet the standards included in the checklist.

Copyright retained by the authors

Theoretical framework

A 360-degree assessment focuses on personnel evaluation. The Joint Committee on Standards for Educational Evaluation has developed standards for personnel evaluation (1988). These standards focus as well on the quality of the evaluation instruments used as on the process of evaluation, and fall into four categories: propriety, utility, feasibility, and accuracy.

The *propriety standards* reflect the fact that personnel evaluations may fail to address, or violate certain ethical and legal principles. They include recommendations that are meant to promote the accessibility of evaluation reports and guidelines for interactions with the person being evaluated. These standards appear to be very important for 360-degree assessment, especially for the assessment process. Ratees as well as raters should be approached with care, since giving and receiving feedback can be threatening. Issues like the purpose of the assessment, or confidentiality should be communicated clearly.

Utility standards are intended to guide evaluations, so that they will be informative, timely, and influential. The evaluation is only informative is the right questions are being asked. Therefore, ratings should focus on competencies that are relevant to the job and should be based on function or task analysis. Items and response wording should also be clear and easy to be understood. Preferably, there is room for recommendations to give raters the opportunity to explain themselves and to provide more useful information. Other important standards for 360-degree assessment are functional reporting and follow-up. Yukl and Lepsinger (1995) developed some guidelines for the display of feedback in the final report. In general, the report should clearly identify feedback from different perspectives, compare feedback from others with the manager's own perceptions, compare the manager's ratings with norms, display feedback for items as well as scales (mean score, range, and distribution), and should provide feedback on recommendations. Follow-up appears to be a crucial factor in a 360-degree assessment (Hazucha et al., 1993). The trainee should be helped to understand the results and pursue appropriate actions.

The *feasibility standards* promote evaluations that are efficient, easy to use, viable in the face of social, political, and governmental forces and constraints, and that will be adequately funded. These standards may be problematic in relation to 360-degree assessment, since this kind of assessment may not be very practical and/or easy to use.

Accuracy standards aim at determining whether an evaluation has produced sound information. Though the other categories mainly focus on the assessment process, the accuracy standards refer especially to the 360-degree instrument. The instrument should be valid, reliable and control bias. Van Velsor et al. (1997) indicate minimal levels of internal consistency, interrater reliability, and test-retest reliability (see checklist).

When studying 360-degree assessment, the performance appraisal literature can also provide valuable input. Performance appraisal concerns the process of identifying, observing, measuring, and developing human performance within organizations (Cardy and Dobbins, 1994). The purpose, in general, is to improve employees' performance and to provide information that can be used in making work-related decisions. Cascio has done important work regarding performance appraisal. According to Cascio (1995) appraisal systems have to meet five key requirements: they should be acceptable, practical, relevant, sensitive and reliable. In the view of Cascio, validity can not be measured directly, since it is unknown what truth' is in performance appraisal. By making appraisal systems relevant, sensitive, and reliable, it can be assumed that the resulting judgments are valid as well. One of the main implications of performance appraisal theory for 360-degree assessment is the importance of using behaviorally focused items in questionnaires (Murphy and Cleveland, 1995). Questionnaires should always be formulated in terms that are easy to interpret by raters. If the assessment is used to measure training transfer, only characteristics that can be controlled and developed by the trainees should be used. To warrant reliability and bias control, raters should be trained to use the instrument properly.

Many authors in the field of 360-degree assessment have established guidelines or recommendations to guarantee proper use (Dalessio, 1998; Tornow and London, 1998; Van Velsor et al., 1997).

On the basis of the aforementioned references, a checklist has been developed for judging the quality of 360-degree instruments (Table 1).

Method

Four instruments for 360-degree assessment have been scored on each element of the checklist. It was attempted to examine all 360-degree instruments developed and sold by Dutch organizations. It is difficult to say how many Dutch instruments are available, but we expect that there are considerably more instruments than the ones that have

been examined here. However, it is supposed that the examined instruments give a good picture of what is available on the Dutch market. The organizations approached to participate in the study delivered a sample of their instrument, a sample of the 360-degree feedback report that ratees receive after the assessment, and all other required information. Additionally, vendors of the instruments were interviewed.

The scoring is based on document analysis (the questionnaires and feedback samples), and on the information provided by the vendor. For some instruments, studies on their reliability and validity are available that provide interesting additional information. If relevant, the results of these studies are presented here.

Table 1
A standards checklist for examining 360-degree instruments

Competencies	Ratings are made on competencies relevant to the job Competencies are based on function or task analysis
Items	Every competency is measured by several items Items focus on behavior that can be observed easily The item wording is clear Only factors over which the ratee has control are included Item wording and qualification are free of irrelevant characteristics, such as race, age, sex, religion
Adaptation	The questionnaire can be adapted to a specific situation
Response	Response scales are clear There is room for recommendations There is a non-response option ('don't know', 'not applicable')
Feedback report	Language and graphics are understandable Feedback is displayed from different perspectives Self-scores are compared to scores of other rater groups Feedback is reported for items as well as competencies The mean score of each item and competency is displayed The range of each item and competency is displayed Scores are compared to norms
Development	The instrument is based on a combination of theory, research and experience
Reliability	All scales have internal consistency coefficients (alpha) of at least .6 Interrater (within-group) reliability is at least .4 All scales have test-retest coefficients greater than .4
Validity	Research should be done to establish validity

Results

In this section the features of the examined instruments are described and their scores with respect to the judgment on the basis of the checklist are presented.

Instrument 1 is a 360-degree instrument which has been available since 1993. It is being used in many Dutch organizations as well as in Japan, Belgium and Great Britain. It is mainly used for management development, though recently a pilot started in which it is used for measuring behavioral change as a result of training. To use this instrument, certification is required and a coaching and follow-up process should be developed. Van der Woude (1995) examined the reliability and validity of this instrument, while Van der Giessen (1997) used it to examine the influence of 360-degree feedback over time on 12 selected behavioral characteristics. The results of these studies can be found in Table 2.

Instrument 2 was developed in 1997 in co-operation with a Dutch university (Rietveld, 1997). The instrument is being used in the Netherlands only. It focuses mostly on high-level managers and is used for developmental, appraisal and communication purposes. The use of the instrument as a training tool is still in the experimental stage. Rietveld (1997) has done some research, concerning the internal consistency and interrater agreement (supervisor versus others and self versus others) of the instrument. Boers (1999) formulated some new competencies and examined their internal consistency. The results of these studies are included in Table 2.

Instrument 3 was developed in 1997 and is being used in many Dutch organizations. Although its main purpose is the development of (higher level) managers and management teams, the instrument is also used in a training context, either before training (to assess the strengths and weaknesses of trainees), or after training, (to

assess further development areas). The intention is to use the instrument for measuring training effects. No research has yet been done on the reliability or validity of this instrument.

Instrument 4 was developed in 1997 and has been used in many Dutch organizations. Free discussion of feedback results is considered the main advantage of 360-degree assessment. The instrument is mainly used prior to training, to assess training needs. Although the instrument now is only sold in combination with training, the intention is to sell it in other contexts than the training context. Groeneveld (1997) examined the internal consistency of the competencies in the instrument. The results can be found in Table 2.

Table 2
The results of the comparison of four 360-degree instruments with the checklist

	Instrument 1	Instrument 2
Competencies	The user (organization/ratee) selects the relevant competencies Competencies are not based on function or task analysis	Ratee and supervisor select relevant competencies Competencies are based on experience with performance appraisal and function analysis
Items	All dimensions contain about 5 items Items are opposite statements about behavior Items focus on easy to be observed behavior Many items measure more than one skill/behavior Items focus on skills that can be developed Item wording does not explicitly contain bias characteristics	All dimensions are measured by 4-8 items Items do not focus on easy to observe behavior Items contain words as: often, sometimes, usually Statements are alternately positively, or negatively formulated Items focus on skills that can be developed Item wording does not explicitly contain bias characteristics
Adaptation	The instrument can not be adapted to a specific situation	The instrument can not be adapted to a specific situation
Response	1-5 scale, position between opposite statements No non-response option No room for recommendations	Response scales are clear 5-point agreement scales Non-response options No room for recommendations
Feedback report	Scores/graphics are explained and understandable Feedback is displayed from different perspectives Self-scores are compared to other-scores Feedback is reported for competencies and items The mean and range of each item and competency is displayed Scores are compared to norms The report includes a strength-weaknesses display (highest/lowest average scores) There are development directions	Graphics are explained and understandable Scores are percentages Feedback is displayed from different perspectives Self-scores are compared to other-scores Feedback is reported for competencies and items Item display is complex because of positively and negatively formulated items There is no mean score or range, since scores are presented as percentages Scores are not compared to norms There are some follow-up directions
Development	Developed on the basis of experience and research	Experience and some research
Reliability	Almost all alpha's are > 0.6 Interrater reliability is low Test-retest is moderate to low	Most alpha's are > .6 Interrater reliability is low Test-retest reliability has not been studied
Validity	Construct validity was studied and considered satisfactory	Validity has not been studied
Additional information		
Experience	six years of experience; been used in many Dutch organizations and in other countries	used for two years in many Dutch organizations
Purpose	Development	Development, communication, appraisal

(Table 2 continued)

	Instrument 3	Instrument 4
Competencies	127 items of which 96 focus on management skills and 31 on personality traits It is a standard instrument (no selection)	39 competencies Ratee and supervisor select relevant competencies Competencies are based on literature research, interviews and expert knowledge
Items	Items do not always focus on easy to observe	Every competency is measured by 5 items

	behavior (personality traits) Not all items can be developed or controlled by the ratee Item wording is clear Item wording does not explicitly contain bias characteristics	Many items do not focus on behavior that is easy to observe, but on personality traits Not all items are easy to be developed by the ratee Item wording is clear Item wording does not explicitly contain bias characteristics
Adaptation	The only adaptation is that items can be left out.	The instrument can not be adapted to a specific situation
Response	Response scales are clear 9-point frequency scale Respondent indicate what the desired score is Non-response option No room for recommendations	Response scales are clear 6-point applicability scale Non-response option Room for recommendations
Report	Feedback is extensively explained Feedback display is complex Feedback is displayed from different perspectives Self-scores are compared to other-scores Mean and range is reported for items and for dimensions Scores are not compared to norms	No display from different perspectives (other-ratings are combined in one group) Self scores are compared to aggregated other scores In the paper-and-pencil version, feedback is only reported for competencies, the electronic version will report on both competencies and items Mean and range for each competency is displayed (in the electronic version also for items) Scores are not compared to norms
Background	Developed on basis of theory	Theory and expert opinions
Reliability	Internal consistency has not been studied Interrater reliability has not been studied Test-retest reliability has not been studied	Internal consistency is low Interrater reliability has not been studied (check) Test-retest reliability has not been studied
Validity	Validity has not been studied	Validity has not been studied
Additional information		
Experience	2 years, used in many organizations	2 years, used in many organizations
Purpose	Development	Input for training

Discussion

In this section each element of the standards checklist is discussed and similarities and differences between the four instruments are described. On the basis of the results presented, organizations and researchers can select the instrument that best meets their needs. For example, an organization that wants to use 360-degree assessment to stimulate communication about managers' job-performance will look for another instrument than a researcher that wants to measure training effects.

Competencies and items

- All instruments except one allow competency selection. Competency selection is necessary when the instrument is being used for training evaluation since only relevant competencies, i.e. those competencies that are developed in a specific training, should be included;
- Every competency usually is covered by several items;
- Though according to the literature it is important that items focus on easily observed behavior, most of the examined instruments contain one or more items that are not formulated into that direction. If items do not focus on observable behavior, raters need to interpret their meaning which probably will result in decreasing reliability. If the instrument is especially meant to be used for development purposes, this may not be a big problem. In contrast, if ratings are being used for training evaluation, reliability is of great importance;
- Some instruments focus on behavior that can be developed by the ratee (instrument 1 and 2) while others contain behavior that can not directly be influenced by the ratee. For example, personality traits are difficult to change and thus not easy to be developed by ratees. Likewise, items formulated as results may be strongly dependent on other factors than the ratee's performance;
- Whether the examined instruments do not cause bias is difficult to say on the basis of the available information. However, none of them does explicitly contain bias characteristics;

- Most instruments are characterized by clear item wording, though some of them contain items that focus on more than one thing at the time, e.g. this person communicates about important decisions *and* asks for input'. These kinds of items can not be answered on one response scale and therefore should be reformulated.

Adaptations

- When the instrument is being used for training evaluation, it should be possible to select the competencies and items on which a specific training focuses. Most instruments do not allow further adaptation than selecting relevant competencies (instruments 1,2,4) or leaving out items (instrument 3). However, in most cases the database of competencies and items is large enough to enable the selection of relevant items.

Response scale

- Two instruments consist of 5-point response scales, one has a 6-point scale, and one a 9-point-scale. If the aim is to measure changes as a result of training this may ask for a, more specific, 9-point scale. However, 5-point scales also seem to provide enough opportunity to measure changes. In addition, most people are familiar with giving ratings on a 5-point scale. A 6-point scale could have the advantage that 'neutral' ratings are not possible since it does not have a central position, such as 5- and 9-point scales. It is difficult to say which response scale is the best option;
- Response scales are formulated as agreement, (for example: 'I agree/I do not agree with this statement'), frequency ('my supervisor often/never exhibits this behavior'), or applicability ('this statement is applicable/not applicable'). Instrument 1 is the only one that has a non-numbered scale where raters must choose between opposite statements;
- All instruments but one (1) have a non-response option;
- Only one instrument (4) contains room for recommendations;
- There is only one instrument (3) where raters also indicate the desired score on each item.

Feedback report

- All instruments, except the fourth instrument, display feedback from different perspectives. Instrument 4 only displays self-scores and aggregated other-scores. For developmental purposes, the feedback should be displayed from different perspectives so that ratees receive relevant information. If the purpose of the instrument usage is training evaluation, as in this case, feedback should also be provided for each different rater source since one of the aims is determining the extent of agreement between and within sources;
- All instruments compare self-scores to other-scores. Again, if the aim is development this is extremely important. This often is the main information of interest to ratees and a starting point for their development;
- All instruments display feedback for both competencies and items. This is important since a score on a competency is not specific enough. For example, if scores for one competency are low, the ratee may want to know which behavior is responsible for it;
- All instruments report the means and range, except instrument 3, which reports results as a percentage. The availability of means and ranges (and the standard deviation) is important if changes between two ratings, (before training and afterwards) are being measured;
- Instrument 1 is the only instrument that compares scores to norms;
- Instrument 1 is the only instrument that contains a strength-weakness display;
- Instrument 1 and 2 contain follow-up directions.

Instrument development

- The basis for the instrument consists usually some sort of experience and/or theory, however no research. Though instruments are used in many organizations, not much is known about their reliability and validity.

Reliability

- Some research concerning their reliability has been done for three of the instruments. Internal consistency appears to be satisfying (.6 or higher) for instrument 1 and 2, but low for instrument 4. Whenever interrater

reliability has been studied it appears to be low. Test-retest reliability has only been studied (in a limited way) for instrument 1 and appeared to be low.

Validity

- Validity has almost never been studied. Instrument 1 is the only instrument for which content validity was studied; it was found to be satisfactory.

Conclusion and limitations

Four Dutch instruments have been evaluated on the basis of a checklist containing important standards. The outcomes of this study give insight into the qualities of a number of 360-degree instruments in terms of their accordance with generally accepted personnel evaluation standards.

The examined instruments do not meet the standard that items should focus on easily observable behavior. Most of the examined instruments contain one or more items that are not formulated that way. Furthermore, only two of the examined instruments focus strictly on behavior that can be developed by the ratee. Regarding the feedback display, most instruments display feedback from different perspectives and all compare self-scores to other-scores and moreover, display feedback for competencies as well as items. The instruments are mainly used for personnel development, however vendors are more and more considering, or experimenting with, the use of 360-degree instruments for other purposes.

Depending on the specific purpose an organization has for a 360-degree instrument, it might prefer one or the other. On the basis of this study it is difficult to indicate one appropriate instrument to be used in training evaluation. Instrument 1 would probably be the most appropriate instrument for a research purpose. This is the only instrument that combines the possibility of competency selection, behavioral focus, focus on items that can be controlled and developed by the ratee, and feedback display of the mean and range for competencies as well as items, and from different perspectives. However, some items may need to be reformulated since they measure more than one behavior.

This study was limited by the fact that only a sample of Dutch instruments was involved. In the future, more instruments should be examined. This study can be used as a starting point for examining other 360-degree instruments.

Furthermore, the examination was based solely on document analysis and interviews with the developers and vendors of the instruments. Thus, the instruments have not been studied in practice and no raters and ratees have been interviewed.

Nevertheless, this paper provides a sound basis for developing better assessment instruments that can be used in the context of human resource development to accomplish various goals, one of them being the measurement of training transfer.

References

Boers, C. (1999) *Voor een spiegel: uitbreiding en aanpassing van een 360-graden feedback instrument.* [Looking in a mirror: expansion and adaptation of a 360-degree feedback instrument]. Thesis, Tilburg

Cardy, R. and G. Dobbins (1994) *Performance appraisal: alternative perspectives*. Cincinnati: South-Western Publishing Co

Cascio, W. (1995) *Managing human resources*. New York: McGraw-hill

Church, A. (1999) *Do higher performing managers actually receive better ratings? A validation of multirater assessment methodology*. Paper presented at the Annual Conference of the Academy of Human Resource Development

Dalessio, A. (1998) Using multisource feedback for employee development and personnel decisions. In: Smither, J.W. (ed.), *Performance Appraisal: state of the art in practice*, San Fransisco: Jossey-Bass

Giesen, R. van der (1997) *Top-performance door 360-graden feedback en coaching* [Top-performance by using 360-degree feedback and coaching]. Thesis, Leiden

Groeneveld, M. (1997) *Competentiegericht opleiden bij klanten van Boertien & Partners*. [Competency-based training of clients of Boertien & Partners]. Thesis, Enschede.

Hazucha, J., S. Hezlett and R. Schneider (1993) The impact of 360-degree feedback on management skills development. *Human Resource Development*, summer/fall, 325-351

The Joint Committee on standards for educational evaluation (1988) *The personnel evaluation standards: how to assess systems for evaluating educators.* Newbury Park: Sage

Lepsinger, R. and A. Lucia (1997) 360-degree feedback and performance appraisal. *Training*, September, 62-70

McLean, G., M. Sytsma and K. Kerwin-Ryberg (1995) *Using 360-degree feedback to evaluate management development: new data, new insights*. Paper presented at the Annual conference of The Academy of Human Resource Development

Morical, K. (1999) A product review: 360 assessments. *Training & Development,* April, 43-47

Murphy, K. and J. Cleveland (1995) *Understanding performance appraisal: social, organizational, and goal-based perspectives*. Thousand Oaks: Sage

Nijhof, W. and A. Jager (1995) *Reliability testing of multirater feedback*. Paper presented at the Annual Conference of the Academy of Human Resource Development

Rietveld, A. (1997) *De 360-graden methode als eye-opener: de ontwikkeling van een feedback instrument*. [360-degree assessment as an eye-opener: the development of a feedback instrument]. Thesis, Tilburg

Robinson, D. and J. Robinson (1989) Training for impact: how to link training to business needs and measure the results, San Fransisco: Jossey-Bass

Rosti, R. and F. Shipper (1998) A study of the impact of training in a management development program based on 360 feedback. *Journal of Managerial Psychology,* no1/2, 77-89

Tornow, W. and M. London (1998) *Maximizing the value of 360-degree feedback*, San Fransisco: Jossey-Bass

Velsor, E. van and J. Leslie (1991) *Feedback to managers: a review and comparison of sixteen multi-rater instruments*. Greensboro: Center for Creative Leadership

Velsor, E. van, J. Leslie and J. Fleenor (1997) *Choosing 360: a guide to evaluating multi-rater feedback instruments for management development*. Greensboro: Center for Creative Leadership

Woude, M. van der (1995) B*eoordelingsprocessen binnen 360-gradenfeedback instrumenten*. [Appraisal processes within 360-degree feedback instruments]. Thesis, Leiden.

Yukl, G. and R. Lepsinger (1995) How to get the most out of 360-degree feedback. *Training*, December, 45-50

BEST COPY AVAILABLE